Enzymatic mechanism of α/β hydrolase fold epoxide hydrolases

The active site of α/β hydrolase fold epoxide hydrolases is situated at the interface of the two essential domains of these proteins, i) the so-called α/β hydrolase fold domain and ii) the lid domain, which is sitting on top of the former. Essentially five amino acid residues are involved in the catalysis. Two tyrosine residues reaching from the lid domain into the substrate binding pocket form hydrogen bonds to the epoxide oxygen of the substrate molecule that enters the pocket. This keeps the substrate in the proper spatial position and also activates it for the first enzymatic step, the nucleophilic attack. The catalytic nucleophile, invariably an aspartic acid residue, forms an ester bond to one of the epoxide ring carbon atoms, while the oxygen-carbon bond in the ring opens under formation of a hydroxy group adjacent to the ester bond. The required proton is provided by one of the tyrosine residues. The second catalytic step, the hydrolysis, is catalyzed by a charge relay system composed of a histidine residue and an acidic residue, again an aspartic acid residue in many EHs. The charge relay system activates a water molecule by proton abstraction that, in turn, hydrolyzes the ester bond between enzyme and substrate through attack as a hydroxyl anion at the carboxylate carbon of the catalytic nucleophile. As a result, the product of the reaction, a so-called vicinal diol, is released and the enzyme is regenerated. For historical reasons, catalytic nucleophile and charge relay system are collectively termed the catalytic triad of the enzyme because the essential function of the tyrosine residues has become obvious only much later.

Sequence analysis of the novel putative epoxide hydrolases

The first sequence comparison of epoxide hydrolases (Arand *et al.* (1994) Sequence Similarity of Mammalian Epoxide Hydrolases to the Bacterial Haloalkane Dehalogenase and Other Related Proteins - Implication for the Potential Catalytic Mechanism of Enzymatic Epoxide Hydrolysis. *FEBS Lett* 338: 251-256) revealed two neighbouring hot spots of sequence conservation, 6 and 5 amino acids in length. Further analysis (Arand *et al.* (1996) in *Control Mechanisms of Carcinogenesis*, eds. Hengstler and Oesch, pp. 116-134) allowed to combine and extend these (underlined) to a single, continuous 16 amino acid motif, <u>RVIAPDLRGYGDSDKP</u>, with an average degree of conservation of 60% per residue. The candidates obtained with the BLASTP search of the human "build protein" database using the above peptide sequence as the bait were further screened for the presence of the following motifs:

- 1) a H-G-X-P tetrapeptide about 15-30 aa N-terminal to the location that matched the bait sequence; this is a conserved sequence, forming part of the oxyanion hole, a structural feature that stabilizes the tetrahedral intermediate formed during the second, hydrolytic step of the catalysis; in epoxide hydrolases, X is usually an aromatic residue, often a tryptophane, while dehalogenases, the closest relatives of EHs, usually have a glutamic acid residue in this position
- 2) a sm-X-D-hy-sm-sm hexapeptide 28-32 aa C-terminal to the location that matched the bait sequence; this is the motif containing the catalytic nucleophile, invariably a D in epoxide hydrolases and dehalogenases, and an S in most other α/β hydrolase fold proteins; sm needs to be a small aa residue, due to sterical constraints in the structure of the catalytic elbow, X is not clearly defined but often a histidine in epoxide hydrolases and hy is usually a hydrophobic amino acid, often a tryptophane in epoxide hydrolases
- 3) two tyrosines in the region of the potential lid domain, in a distance of 50-85 aa from each other; unfortunately, the sequence context of the first tyrosine is not at all conserved; the second one is located in a tripeptide with the consensus sequence Y-R-N, where R may be replaced by K, and N may be replaced by D or E, but other substitutions occur as well; these two tyrosines are involved in the initial substrate recognition and activation of the oxiran ring and are absent in dehalogenases and esterases
- 4) a poorly conserved tetrapeptide G-E-L-D, roughly 30 aa C-terminal to the second tyrosine; the D (or E in case of mEH-related enzymes) is the acidic residue of the charge relay system that catalyzes the hydrolysis of the covalent intermediate, the G is the second best conserved residue within this motif; this motif is not conserved in mEH-related EHs, and D is replaced by E; in some microbial EHs, the localization of the acidic residue in the amino acid sequence is apparently not conserved
- 5) a moderately conserved heptapetide G-H-W-T-Q-I-E, ≥ 20 aa upstream of the C-terminus; H is the calatytic histidine, affording the water activation in the charge relay system

The position of these motifs is indicated in the sequence comparison below (Fig. SI2), that also forms the basis for the phylogenetic tree in Fig. 1 of the manuscript, as well as for the selection of candidate catalytic residues that were mutagenized.

	10	20	30	40	50	60	70	80
eh3								0
Н	***************************************							0
1/MEST	MALON STREET DOUG	AT DAVEGUE C	DECENT AT DO	TI I NO TO VOC	DECAMENT M	COTTO COMT	or uppyioned	SETAKV 8
3	MTLRAAVFDLDGVLALPAVFGVLGRTEEALALPRGLLNDAFQKGGPEGATTRLMKGEITLSQWIPLMBENCRKCSETAKV							
4								0
h1 h2					********			0
onsensus	XXX	KLXLG	FARLWLMLRX	CX	WG			
	90	100	110	120	130	140	150	160
h3	MGILKLI							
H 1/MEST				CEETLPLEDGW				IR 5
1	CLPKNFSIKEIFDKA							
4				VE				
h1				TCNCP-NYKI				
h2				P				
nsensus		xxx	DLXORXXXXX	LGLLTEXXXD	XXXGFXXXOI	SXXPACLXN	K	-XWSHX
	170	180	190	200	210	220	230	240
h3	PF)					The second second	7.4	-
н				TP-PLEDS-C				
1/MEST				VGLLAVPLLA				
3	PEPQIYKFLLDTLK							
4		LVY	CYCGLCASIH	LIKLLWSLGKG	PAQTFRRPAR	REHPPACLSDI	PS	-LGTHC 7
h1 h2				SAILFLGFAV				
			QIIITIGA	LLTLTWKWET	EGNETEVERY	PEPECLEN.		WNHK 5
nsensus	XVTLKX	-XLRLHYVXX	GPXX	CPLMLLLHGFP	EXWYSWRXQI	PEXTD	XXEV	RKCKAL
	250	260	270	280	290	300	310	320
h3	EATINOFPOFKTEIR	EGLQVHFLHV	KPPKSYKN	KPILVAHGWP	GNVFEFYKF	PLLTDPKKH	GIDSDFAFEV	IAPSIP 1
Н	VEILNRYPHFKTKIE							
1/MEST	FFTYKG							
3	FLNLKSS	GLRLHYVSA	GRGN	PLMLFLHGFP	ENWFSWRYQI	REFQS	RFHV	VAVDLR 1
4 eh1	YVRIKDS	CTPEUVIA	ORD OIL	PRICES TO THE PROPERTY	PERVONDAVAL	REFES	EYRV	VALDLR 1
								m Thon 1
	YIKLKK	VRLHYVQT	GSDDI	CPLMLFIHGYP	EFWYSWRFQI	KEFAD	KYRC	
eh2	YIKLKK	IRMHYVEE	GSDDI GPADO	CPLMLFIHGYP GDVLLMVHGFP	efwyswrfqi efwyswrfqi	KEFAD LEHFKH	KYRC	IAIDMR 1
h2	FVQLKNGYGXSDAPX-XDXYS	VRLHYVQT IRMHYVEE SIDXLARDIR	GSDDI GPADO DLXLXLGYS-	KPLMLFIHGYP SDVLLMVHGFP -KXXLXGIQVG	EFWYSWRFQI EFWYSWRFQI	YPEXVXXLX	KYRC	IAIDMR 1
eh2 onsensus	YIKLKKFVQLKNGYGXSDAPX-XDXYS	VRLHYVQT IRMHYVEE SIDXLARDIR 340	GSDDI GPADO DLXLXLGYS 350	CPLMLFIHGYP SDVLLMVHGFP -KXXLXG DVG 360	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370	LEHFKH LYPEXVXXLX: 380	KYRC THRC KXNXPHPSXF	IAIDMR 1 XEXXLX 400
eh2 onsensus eh3	YIKLKKFVQLKN	-VRLHYVOTO -IRMHYVEE SIDXLARDIR 340 SQLACARVER	GSDDI GPADI DLXLXLGYS- 350 KLMLRLGYN	CPLMLFIHGYP SDVLLMVHGFP -KXXLXG DWG 360 -KFYLQGGDWG	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKV	LEHFKH LEHFKH LYPEXVXXLX 380 VYPQNVMALH	XXNXPHPSXF: 390 LNMMPAMPGAI	XEXXLX 400 NALGTF 2
eh2 onsensus eh3 EH g1/MEST	YIKLKK	VRLHYVQT IRMHYVEE SIDXLARDIR 340 SQLACARVFR NSVATARIFY SIFEQASIVE	GSD	CPLMLFIHGYP CDVLLMVHGFP -KXXLXGLDGG 360 -KFYLQGGDWG EFYLQGGDWG RRINLLSHDYG	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI DIVAQELLYF	KEFAD EHFKH 380 YYPONYMALHI VPSHVKGLHI KYKONRSGRL		XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1
eh2 onsensus eh3 EH g1/MEST H	YIKLKK- FVQLKN	VRLHYVQTIRMHYVEE SIDXLARDIR 340 SQLACARVFR NSVATARIFY SIFEQASIVE CMEVLCKEMV	GSD	CPLMLFIHGYP CDVLLMVHGFP -KXXLXGLDGG 360 - -KFYLQGGDWG -KFYLQGGDWG -EFYIQGGDWG RRINLLSHDYG -QAVFIGHDWG	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKV SLICTNMAQI DIVAQELLYF GMLVWYMALF	KEFAD EHFKH 380 YYPQNVMALHI VYPSHVKGLHI KYKQNRSGRL FYPERVRAVAS	KXNXPHPSXF 390 LINMMPAMPGAI LINMALVLSNF FIKSLCLSNG SLNTPFIPAN	ALIDMR 1 XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 PNMSPL 3
eh2 onsensus eh3 EH	YIKLKK	-VRLHYVOT -IRMHYVEE SIDXLARDIR 340 SQLACARVFR NSVATARIF SIFEQASIVE CMEVLCKEMV	GSD	RPLMLFIHGYP BOVLIMVHGFP -KXXLXG D -KFYLQGGDWG -EFYLQGGDWG -EFYLQGGDWG -EFYLGGDWG -GAVFIGHDWG -KCILVAHDWG	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKQ SLICTNMAQI DIVAQELLYF GMLVWYMALE ALLAWHFSIX	LEFAD LEFEKH .YPEXVXXLX: 380 YPONVMALHI LVPSHVKGLHI YYKONRSGRLI TYPERVRAVA:	KXNXPHPSXF	AUDMR 1 AUD
eh2 onsensus eh3 H g1/MEST H 3 4	YIKLKR-FVQLKN	-VRLHYVQT -IRMHYVES SIDXLARDIR 340 SQLACARVFR: NSVATARIFY: SIFEQASIVE: MEVLCKEMV: TIVLLEVDIK: KLDCLITDIK: SIDELTGDIR:	GSDDI GPADI DLXLXLGYS 350 KLMLRLGYN KLMIRLGFQ ALLRHLGLQNI TFLDKLGLS DVLLGLGYS DVLIGLGYS DVLIGLGYS	CPLMLFIHGYP COVILMVHGFP STOUL STOU	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI DIVAQELLYF GMLVAWHFAII GMLWWYMALE GMLWWYFAEG GMLWWJFAEG GMLWWQFAEG GLVAWQFAEG	LEHFKH		ALGUMR 1 XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 PNMSPL 3 QDYSLH 2 TEYILR 2 RKRIYT 2
eh2 onsensus eh3 H g1/MEST H 3 4	YIKLKR- FVQLKN	-VRLHYVQT -IRMHYVES SIDXLARDIR 340 SQLACARVFR: NSVATARIFY: SIFEQASIVE: MEVLCKEMV: TIVLLEVDIK: KLDCLITDIK: SIDELTGDIR:	GSDDI GPADI DLXLXLGYS 350 KLMLRLGYN KLMIRLGFQ ALLRHLGLQNI TFLDKLGLS DVLLGLGYS DVLIGLGYS DVLIGLGYS	CPLMLFIHGYP COVILMVHGFP STOUL STOU	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI DIVAQELLYF GMLVAWHFAII GMLWWYMALE GMLWWYFAEG GMLWWJFAEG GMLWWQFAEG GLVAWQFAEG	LEHFKH		ALGUMR 1 XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 PNMSPL 3 QDYSLH 2 TEYILR 2 RKRIYT 2
eh3 H H Bg1/MEST H 3 4 4 bh1 bh2	YIKLKR-FVQLKN	-VRLHYVOT -IRMHYVE SIDXLARDIR 340 SQLACARVPR SSVATARIFY SIFEQASIVE CMEVLCKEMV TIDLLLVDIK KLDCLITDIK SIDELTGDIR RLTHLVEDIR	GSDDi GPADi DLXLXLGYS 350 KLMLRLGYN ALLRHLGLQNI TFLDKLGLS DVILGLGYS DVILGLGYS DVILGLGYS DVIEGLGYS DVIEGLGYD QFIEILELK	CPLMLFIHGYP COVILMVHGFP STOUL STOU	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTK SLICTNMAQI DIVAQELLYI GMLVWYMALI ALLAWHFSI GMIAWLIAIC GLVAWQFABC AIVCWRVAMI	MEFAD		AUDINE 1 XEXXLX 400 NALGTF 2 STLTIL 2 GIFPET 1 PNMSPL 3 QDYSLH 2 TEYILR 2 RKRIYT 2 FEVYNM 1
eh2 ensensus eh3 EH gg1/MEST H 3 4 eh1 eh2	YIKLKR-FVQLKN	-VRLHYVOT -IRMHYVE SIDXLARDIR 340 SQLACARVPR SSVATARIFY SIFEQASIVE CMEVLCKEMV TIDLLLVDIK KLDCLITDIK SIDELTGDIR RLTHLVEDIR	GSDDi GPADi DLXLXLGYS 350 KLMLRLGYN ALLRHLGLQNI TFLDKLGLS DVILGLGYS DVILGLGYS DVILGLGYS DVIEGLGYS DVIEGLGYD QFIEILELK	CPLMLFIHGYP DVLLMVHGFP - KXXLXGLOWG 350 - KFYLQGGDWG - KFYLQGGDWG - EFYLQGGDWG - EFYLQGGDWG - CPYLGGBWG - CPYLGGBWG - CPYLGGBWG - KCILVAHDWG - KCYLIGBBWG - KCYLIGBBWG - KAIVYAHDWG - RVTLAAHDWG	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTK SLICTNMAQI DIVAQELLYI GMLVWYMALI ALLAWHFSI GMIAWLIAIC GLVAWQFABC AIVCWRVAMI	MEFAD		AUDINE 1 XEXXLX 400 NALGTF 2 STLTIL 2 GIFPET 1 PNMSPL 3 QDYSLH 2 TEYILR 2 RKRIYT 2 FEVYNM 1
ph2 ph3 ph3 ph3 ph3 ph3 ph3 ph4 ph4 ph1 ph2 ph2 phsensus	YIKLKR-FVQLKN	-VRLHYVOT -IRMHYVEE SIDXLARDIR 340 SQLACARVFR SVATARIFY SIFEQASIVE MEVLCKEMV IDLLLVDIK KLDCLITDIK SIDELTGDIR NLTHLVEDIR VLCQU 420 SKEIQKTHNP	GSDDi GPADi GPADi DLXLXLGYS 350 KLMLRLGYN KLMRLGFQ ALLRHLGLONI TFLDKLGLS DVILGLGYS DVILGLGYS DVILGLGYS DVILGLGYD QFIEILELK FXKSXVXLXFY 400 FSK-FGLLIVI	CPLMLFIHGYP DVLLMVHGFP -KXXLXGLOWS 360 -KFYLQGGDWG -KFYLQGGDWG -KFYLQGGDWG -KFYLQGGDWG -KYLLYAHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KYLIAHDWG -RVTLAHDWG -RTGYHHLQATK	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI GIVAWHALE ALLAWHFSIY GMIAWLIAIG GIVAWGFAEG AIVCWRVAMI PX 450 PDTAGTSLNI	MEFAD- LEHFKH		XEXXLX 400 NALGTF 2 STITLL 2 GIFPET 1 FMMSPL 3 QUYSLH 2 TEYILR 2 TEYILR 2 FEVYNM 1 GIRXX- 480 ENRAL- 3
oh3 H 11/MEST H 3 4 bh1 oh2 msensus	YIKLKK- FVQLKN	VRLHYVOT IRMHYVE SIDXLARDIR 340 SQLACARVFR SVATARIFY SIFEQASIVE MEVICKEMV ITDILLVDIK KIDCLITDIK SIDELITDIR NITHLVEDIR 420 SKEIQKTENP ERDVELLYPV	GSD	CPLMLFIHGYP DVLLMVHGFP -KXXLXG DTS 360 -KFYLQGGDWG -KFYLQGGDWG -KFYLQGGDWG -KFYLQGGDWG -KYLLYAHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KCYLIGHDWG -KYLIAHDWG -KYLIAHDWG -RVTLAHDWG -RV	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKV SLICTNMAQI DIVAQELLYF GMLWYMALF ALLAWHFSIY GMIAWLIAIC GIVAWQFAEG AIVCWRVAMI PX 450 PDTAGTSLNIF PDTVGSALNI	MEFAD- LEHFKH		AUDINE 1 AUD
ch2 ch3 ch3 ch4 ch7 ch4 do ch1 ch2 ch3 do ch1 ch2 ch3 ch4 ch1 ch2 ch3 ch3 ch4	YIKLKR-FVQLKN-STARK-STAR	VRLHYVOT IRMHYVEE 340 340 SQLACARVER SVATARIFY SIPEQASIVE MEVICKEMV TIDLLLVDIK KIDCLITDIK SIDELTGDIR NITHLVEDIR 420 SKEIQKTHNP ERDVELLYPV	GSDDi GPADi GPADi DLXLXLGYS 350 KLMLRLGYN ALLRHLGLQNI TFLDKLGLS DVILGLGYS DVILGSLGYS DVILGLGYS CFILLIVI FXKSXYKLXFY GFSK-FGLLIVI EKKVFYSLMRI LQKLLK-I	CPLMLFIHGYP DVLLMVHGFP RXXLXGLOWG 350 -KFYLQGDWG EFY1QGDWG EFY1QGDWG EFY1QGBWG CFY1LQGBWG KCILVAHDWG KCILVAHDWG KCILVAHDWG KCYLIGBBWG KCYLIGBBWG KAIVYAHDWG EVYLAAHDWG DXYYPEXLTS 440 ETGYHHLQATK ESGYMHLQATK ESGYMHLQATK	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SKLCTNMAQI DIVAQELLYY GMLWYMALE ALLAWHFSI) GMTAWLIAII GGUAWQFAE AIVCWRVAMI PX	MEFAD- LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LYPEXVXXLX LYPEXVXLX LYPEXVKALI LYPEXVLX LYPEXVXLX LYPEXVXLX LHSNLIDRLY		ALIDMR 1 XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 PHMSPL 3 QOYSLH 2 TEYLLR 2 TEYLLR 2 TEYLR 2 TEYLR 480 ENRAL 3 ENRAL 3 ENRAL 3 ENRAL 3 ENRAL 3
nsensus th3 H 11/MEST H 3 4 bh1 ch2 nsensus th3 H 11/MEST H 11/MEST H 3	YIKLKK- FVQLKN	VRLHYVOT IRMHYVE SIDXLARDIR 340 SQLACARVFR SVATARIFY SIFEQASIVE MEVICKEMV ITDLLLVDIK KIDCLITDIK SIDELITDIK NITHLVEDIR 420 SKEIQKTENP ERDVELLYPV	GSDDi GPADi GPA	CPLMLFIHGYP DVLLMVHGFP -KXXLXG DTS 360 -KFYLQGGDWG -KFYLQGGDWG -EFYTQGGDWG -KYLLYAHDWG -KCILVAHDWG -KCILVAHDWG -KVLIGHDWG -KVLIGHDWG -KVLIGHDWG -KVLIGHDWG -KYLAHDWG -	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTK\ SLICTNMAQI DIVAQELLYF GMLWYMALE ALLAWHFSI GGIAWLIAIC GIVAWGFAE AIVCWRVAMI PX 450 PDTAGTSLNI PDTVGSALNI LMN MS	MEFAD- LEHFKH		AUDINE 1 AUD
ch3 H g1/MEST H sh4 ch1 ch2 H sh3 H g1/MEST H sh3 H sh4 ch1 ch2 H sh3 H g1/MEST H	YIKLKR- FVQLKN- GYGKSDAPX-XDXYS 330 GYGKSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGFSDAPRDVDCYT GYGETDAPIRE,NYM GYNTIDRPSGISDYP H- 410 Y-DILGMLIPSTLSS LGQRFGRFLG-LTE H- ES-	VRLHYVOT IRMHYVEE 340 340 SQLACARVER SVATARIFY SIFEQASIVE CMEVLCKEMV IDLLLVDIK MLDCLITDIK SIDELTGDIR NLTHLVEDIR 420 SKEIQKTENP ERDVELLYPV IKAI IKAI PAQ PAQ	GSDDi GPADi GPADi DLXLXLGYS 350 KLMLRLGYN KLMLRGQYD DVILGGGYS DVILGSLGYS DVILGSLGYS DVILGLS DVILGLS FXKSXYLLXFY 400 FSK-FGLLIVI KEKVFYSLMRI LQKLLK-I NPVFDYQLYFY EFRSHYMFLFY LLKSSYYYFFY	CPLMLFIHGYP DVLLMVHGFP - KXXLXGLOWG 360 KFYLQGGUWG KFYLQGGUWG KFYLQGGUWG KFYLQGGUWG KYYLQGGUWG KYYLQGGUWG KYYLQGGUWG KYYLQGGUWG KYYLQGGUWG KYYLQGGUWG KYYLQGWUG KY	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY STATE STAT	MEPAD- LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH LEHFKH	- KYRC - THRC 390 LNMMPAMPGAI LNMALVLSNE; IKSLCLSNE; SLNTPF1PAN; VVSGAPMSVV; UNFPBPNVF- CONIPRESSF; LCNVPHPFAF; 470 LEKFSTWTNT; LEKFSTWTNT; SSRGLTEVFG; LSMHRVCEAG; LKHTLTHRKT;	XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 PMMSPL 3 QDYSLH 2 TEYLIR 2 TEYLIR 2 GIRXX- 480 ERRAL 3 EFRYL 3 PYTR- 2 GIFYNS 4 GI
oh3 H 1/MEST H 3 4 H 4 H 1/MEST H 3 H 1/MEST H 1	YIKLKK- FVQLKN	VRLHYVOT IRMHYVEE SIDXLARDIR 340 SQLACARVER SVATARIFY SIFEQASIVE CMEVICKEMV IDLLLVDIK SIDELTGDIR NLTHLVEDIR 420 SKEIQKTHNP ERDVELLYPU	GSDDi GPADi GPA	CPLMLFIHGYP DVLLMVHGFP - KXXLXGLOWS 360 - KFYLQGGWG - KFYLQGGWG - EFYTQGGWG - CAVFIGHDWG - KCILVAHDWG - KCVLIGBDWG - KCVLIGBDWG - KCVLIGBDWG - KCVLIGBDWG - KCVLIGBDWG - KAIVVAHDWG - CVIIGBDWG - CVIIGBDWG - KAIVVAHDWG - CVIIGBDWG - CVIIGB	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI GIVAWGALLY GMIVWYMALE ALLAWHFSII GLVAWQFAEG AIVCWRVAMI PX	MEFAD- LEHFKH	-KYRC -THRC XXNXPHPSXF: 390 LNMMPAMPGAI LMMALVLSNF: ITKSLCLSNG SLNTPFIPAN: VVSGAPMSYY; LNFPEPNVF: CCNIPRPGSF! LCNVPHPFAF: LEKFFTWTXT: 470 LEKFSTWTNT: EFKFSTWTNT: EFKFSTWTTT EFKFSTLTPVFG; LSMHKVCEAG LKTTLTHRKT* LKHLFTSHST* LEKLFTSHST*	AUDIN 1 AUD
nsensus hith Hith Insensus A th A th Insensus	YIKLKR- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGFSDAPRDVDCYT GYMLSDKPKHVDNYS GYNTIDRPSGISDYP 410 Y-DILGWLIPSTLSS LGQRFGRFLG-LTF H	VRLHYVOT IRMHYVEE 340 SQLACARVFR SVATARIFY SIFEQASIVE CMEVICKEMV IDLILVDIK KIDCLITDIK SIDELTGDIR ILTHIVEDIR	GSDDi GPADi GPADi GPADi DLXLXLGYS 350 KLMLRLGYN KLMLRLGYN KLMLRLGYN DLDSLGYS DVIEGLGYS DVIEGLGYD DVIEGLGYD DVIEGLGYD GFSK-FGLLIVI KEKVFYSLMRI LOKLLK-I NPVFDYQLYF FFRSHYMFLF LLKSYYYFFF FFRSWYMFFY RNKSWYIYFF	CPLMLFIHGYP DVLLMVHGFP - KXXLXGLOWG 360 KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KALVYAHDWG CAVFIGHDWG KALVYAHDWG KALVY	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY STATE GHLVWYMALF ALLAWHFSTY GMIZWRYMALF ALLAWHFSTY 450 PDTAGTSLNI PDTVGSALNI LMN QN-LSRTFKS MS MS AD SN-	MEPAD- LEHFKH	- KYRC - THRC 390 LNMMPAMPGAI LNMALVLSNF; IKRSLCLSNG SLNTPFIPAN; VVSGAPMSVV; UNFPEPNVF; CCNIPRPGSF; LCNVPHPFAF; 470 LEKFSTWINT; LEKFSTWINT; LEKFSTWINT; SERGLITVFG; LSMHKVCEAG; LKHLITHSHST; LELCFRAKEI; LEAMFRGSKA	AUDINE 1 AUD
nsensus hith Hith Insensus A th A th Insensus	YIKLKR- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGPSDAPRDVDCYT GYGEDAPIBRONY GYNTDRPSGISDYN H	VRLHYVOT IRMHYVEE 340 340 SQLACARVER SVATARIFY SIPEQASIVE MEVLCKEMV TIDLLLVDIK KLDCLITDIK SIDELTGDIR NITHLVEDIR 420 420 SKEIQKTENP ERDVELLYPV - PAQ - PAQ - WSQ KEQ LXAWVYXFSQ	GSDDi GPADi GPADi GPADi DLXLXLGYS 350 KLMLRLGYN KLMLRLGYN BYLLGLGYS DVILGLGYS DVILGLGYS DVILGLGYS DVILGLGYS DVILGLGYS FXKSXYLLXFY GFSK-FGLLIVI KEKVFYSLMRI LOKLIK-I NPVFDYOLYFY FFRSHYMFLFY LLKSSYYYFFY FRKSWYHFLY RNKSWYHFLY RNKSWYHYLFY PGXLXG-PLN	CPLMLFIHGYP DVLLMVHGFP 350 -KFYLQGDWG -KFYLQGDWG -KFYLQGDWG -KFYLQGDWG -KFYLQGDWG -KFYLQGDWG -KFYLQGBWG -KFYLQGBWG -KFYLQGBWG -KFYLGBBWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KYLAAHDWG -KYLAAHDW	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SSLICTNMAQI DIVAQELLYY GMLWYMALE ALLAWHFSI) GMTAWLIAI GGUAWQFAE AIVCWRVAMI PX	MEPAD- LEHFKH LYPEXVXXLXI 380 YPPQNVMALHI LYKQNRSGRL* YPPERVRAVA: LYPEWMOKLI* LYPEWMOKLI* LHSNLIDRLV* L	KYRC THRC 390 LNMMPAMPGAI LNMALVLSNE ITKSLCLSNE SLNTPFIPAN VINFPBPNVE CONIPRESSF LONVPHPFAF 470 LEKFSTWTNT LEKFSTWTNT SERGLITVFE LSMHRVCEAGE LSMHRVCEAGE LKTLITHRKT LKHLFTSHST LELAFFREE LEAFFREE LEAFFREE LEAFFREE LEAFFRES LEAFRES	ALIDMR 1 AMERICAN AUGUST 2 STLTIL 2 GIFPET 1 FEWNINEL 3 QUYSLH 2 TEYILR 2 TEYILR 2 TEYILR 2 TEYILR 2 TEYILR 2 TEYIN 4 ABO ENRAL 3 ETRYL 3 ETRYL 2 GIFYL 2 GIFYL 2 GIFYL 3 GIFYL 2 GIFYL 3 GIFYL 2 GIFYL 3 GIFYL 2 GIFYL 3 GI
nsensus hith Hith Insensus A th A th Insensus	YIKLKR- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGESSAPPEIEEY GYGESSAPPEIEEY GYGETDAPINRONY GYNTIDRPSGISDY H- 410 Y-DILGWLIPSTLSS LGQRFGRFLG-LTF H- ES- S NFTDEDI 490	VRLHYVOT IRMHYVEE 340 SQLACARVFR SVATARIFY SIFEQASIVE CMEVLCKEMV IDLLLVDIK KLDCLITDIK SIDELTGDIR ILTHLVEDIR	GSDDi GPADi GPA	CPLMLFIHGYP DVLLMVHGFP - RXXLXGLOWS 360 - KFYLQGGWG - KFYLQGGWG - KFYLQGGWG - KFYLQGGWG - KFYLQGGWG - KYLLYAHDWG - QAVFIGHDWG - KCULIGBWG - KAIVVAHDWG - KYLAHUMG - KYL	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY STATE GIVAWGALLY GMLVWYMALH ALLAWHFSIY GMYAWLIAII GLVAWQFAEG AIVCWRVAMI PX	JEPAD- LEHFKH	XXXXPRPSXF: 390 LNMMPAMPGAI LNMALVLSNF; ITKSLCLSNG; SLNTPFIPAN; VVSGAPMSVY; UNFPEPNVF; CCNIPRPGSF; LCNVPHPFAF; 470 LEKFSTWINT; LEKFLFTHST; LELCFRAKE; LEAMFGSKA; DAXLXLETA;	AUDIME 1 AUD
nsensus his	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGESSAPPEIEEY GYGESDAPRDVDCYI GYNLSDKPKHVDNYS GYNTTDRPSGISDYN 410 Y-DILGWLIPSTLSS LGGRFGRFLG-LTI H- ES	VRLHYVOT IRMHYVEE 340 SQLACARVER SVATARIFY STEPQASIVE CMEVICKEMV IDLILIVDIK SIDELIGDIR ALDCLITUDIK SIDELIGDIR ALDCLITUDIK SIDELIGDIR ALDCLITUDIK INLTHIVEDIR	GSDDi GPADi GPA	CPLMLFIHGYP DOVLLMVHGFP	EFWYSWRFQI EFWYSWRFQI XXXAWLXAII 370 AIITSLLTKY SLICTNMAQI GIVAWJMALE ALLAWHFSII GLVAWQFAEG AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LNMMPAMPGAI LMMALVLSNF; TIKSLCLSNG SLNTPFIPAN; VVSGAPMSYY; LNFPEPNVF; CCNIPRPGSF; LCNVPHPFAF; LEKFSTWTNT; EKFSTWTNT; EKHLFTSHST; LEAMFRGSKAG DAXLXLETA; 550 SLNELYDRTP.	ALIDMR A
nsensus h3 H g1/MEST H 3 4 hb1 nsensus h3 H g1/MEST H 3 4 hb1 nsensus h3 H g1/MEST H 3 H h1 h1 H H H H H H H H H H H H H H H H	YIKLKR- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEYC GYGESSAPPELEEYC GYGESSAPPELEEYC GYGESTDAPIRBONYJ GYNLSDKPKHUDNYS GYNTTDRPSGISDYN H	VRLHYVOT IRMHYVEE 340 SQLACARVFR SIDXLARDIRI 340 SQLACARVFR SIFEQASIVE MEVICKEMV TIDLLLVDIK KIDCLITDIK SIDELTGDIRI NITHLVEDIR 420 SKEIQKTHNP ERDVELLYPV	GSD	CPLMLFIHGYP DOVLLMVHGFP RXXLXGLOGG 350 -KFYLQGGDWG -EFYLQGGDWG -EFYLQGGDWG -RVTLQGGDWG -RVTLAHDWG -KCILVAHDWG -	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTK SLICTNMAQI DIVAQELLYI GMLWYMALE ALLAWHFSI) ALLAWHFSI) FOR THE STATE OF T	MEFAD MEFKH MY PEXVXXLX MO MY PONYMALH MY SHV KG HI MY KONSGRL MY PONYMALY MY PERVAVA MACHINE MACHINE MI MACHINE MACH	- KYRC - THRC - THRC - 390 - LNMMPAMPGAI - LNMALVLSNF: - ITKSLCLSNG: - SLNTPF IPAN - LYSEAPMSVY: - LYSEAPMSVY: - LYSEAPMSVY: - LEKFFTWITT: - LEKFFTWITT: - LEKFSTWINT: - EKFSTWINT: - LEKFSTWINT: - STOLL STANLING - LELETTRITT -	ALIDMR 1 AMERICAN AUGUST 2 STLTIL 3 STLTIL
nsensus his H pl/MEST 4 sh his his H his his his H his his H his	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGGSDAPRDVDCYI GYMLSDKPKHVDNYS GYNTIDRPSGISDYN H	VRLHYVOT IRMHYVEE SIDXLARDIR 340 SQLACARVFR SVATARIFY SIFEQASIVE CMEVICKEMV IDLLLVDIK SIDELTGDIR KLDCLITDIK SIDELTGDIR KLDCLITDIK SIDELTGDIR ILTHLVEDIR	GSDDi GPADi GPA	CPLMLFIHGYP DOVLLMVHGFP	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY SLICTNMAQI GIVAWJMALE GLYAWJFSI GKIVWYMALE ALLAWHFSI 450 PDTAGTSLNI PDTYGSALNI LINN QN-LSRTFKS MS IN AD -CEALGKRY- MYCKLERMY MYCKLERMY -WACKSLGRR	MEFAD- LEHFKH LYPENVXLIX LYPENVALH LYPENVALH LYPENVALH LYPENVAKLI LYPENVAKL	XYRC THRC XXNXPHPSXF: 390 LNMMPAMPGAI LMMALVLSNF: TIRSLCLSNG SLNTPFIPAN VVSGAPMSYYV LONUPHPPAF: LEKFSTWTNT: EKFSTWTNT: STRGLTPYGG SLMLYCEAG LATLLEAKT: LEAMFRGSKAG DAXLXLETAI 550 SLNLYDRTP. FFFELLH-TPI: LDPVNPYPEF: AEIONVLVPQI A	ALIDMR A
sh2 msensus sh3 H g1/MEST H 3 4 sh1 sh2 msensus sh3 H 1/MEST H 3 1/MEST H 1	YIKLKR- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGESSAPPELEEY GYGESTDAPIRBONY GYNLSDKPKHVDNY GYNLTDRPSGISDY H	VRLHYVOT IRMHYVEE 340 SQLACARVER STEQASIVE MEVICKEMV TIDLLLVDIK KLDCLITDIK SIDELTGDIR NITHLVEDIR 420 SKEIQKTHNP ERDVELLYPV	GSD	CPLMLFIHGYP DOVLLMVHGFP - EXXLXGLOG 350	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTK SLICTNMAQI DIVAQELLYI GMLWYMALL ALLAWHFSI) GLYAWQFAEG AIVCWRVAMI PX	MEFAD MEFKH MY PEXVXXLX MO MY PONYMALHI MY SHVKGINI	- KYRC - THRC 390 LNMMPAMPGAI LNMALVLSNF: ITKSLCLSNG: SLNTPFIPAN VYSGAPMSVY; VINFPBPNVF CONIPRESSF! ICNVPHPPAF: LEKFSTWINT: LEKFSTWINT: LEKFSTWINT: LEKFSTWINT: LEKFSTWINT: LEKFSTWINT: LEKFSTWINT: SSGLTPVGG LSMHKVCEAGE LELCHFRESKEI LECHFRESKEI LELCHFRESKEI LELCHFRESKEI LELCHFRESKEI LELCHFRESKEI LECHFRESKEI LELCHFRESKEI LELCHFRESKEI LELCHFRESKEI LELCHFRESKEI LECHFRESKEI LEC	ALIDMR 1
sh3 H g1/MEST H 3 4 sh1 sh2 msensus sh3 H g1/MEST H 4 sh1 sh2 msensus sh3 H g1/MEST H 3 4 sh1 H g1/MEST H 3 4 sh1 sh2 msensus	YIKLKR- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFF GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGESSAPPEIEEY GYGESSAPPEIEEY GYGETDAPIHRONY GYNTIDRPSGISDY H- 410 Y-DILGWLIPSTLSS LGQRFGRFLG-LTF H- ES	VRLHYVQT IRMHYVEE 340 SQLACARVFR SVATARIFY SIPEVASIVE CMEVLCKEMV IDLLLVDIK KLDCLITDIK SIDELTGDIR KLDCLITDIK SIDELTGDIR LLTHVEDIR	GSDDi GPADi GPA	CPLMLFIHGYP DOVLLMVHGFP - RXXLXGLOWS 360 RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RYFIGHOWG RYFIGHOWG RYFIGHOWG RAIVAHDWG RAIVAHD RAIVAHDWG	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY STATE GILVEYMALE GLIVAYMALE ALLAWHFSIY GKIVWYMALE ALLAWHFSIY ASTO FOR THE STATE AND THE STATE ST	MEPAD- LEHFKH	XXXXPRPSXF: 390 LNMMPAMPGAI LNMALVLSNF; ITKSLCLSNG SLNTPFIPAN; VVSGAPMSVY; UNFPEPRVF; CCNIPRPGSF; LCNVPHFFAF; 470 LEKFSTWINT; LELGFRAKE; LEAMFGSKA; LEAMFGSK	ALIDMR 1
sh2 msensus sh3 H g1/MEST H 3 4 bh1 sh2 msensus sh3 H g1/MEST H 3 4 4 sh1 sh2 msensus	YIKLKR- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGESSAPPELEEY GYGESTDAPIRBONY GYNLSDKPKHVDNY GYNLTDRPSGISDY H	VRLHYVOT IRMHYVEE 340 SQLACARVER SYLATARIFY SYLATARIFY SYLEPANIVE MEVICKEMV IDLILVDIK SIDELTGDIR KLDCLITDIK SIDELTGDIR KLDCLITDIK SIDELTGDIR ILTHLVEDIR	GSDDi GPADi GPA	CPLMLFIHGYP DOVLLMVHGFP	EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY SLICTNMAQI GIVAWJMALE ALLAWHFSII 450 PDTAGTSLNI PDTYGSALNI LINN	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFF LETONVPLEGGIV. ADGALD LEAA'	ALIDMR A
sh2 msensus sh3 H g1/MEST H 3 4 sh1 sh2 msensus sh3 H g1/MEST H 3 4 sh1 H 3 1 H g1/MEST H 3 3 4 sh1 sh2 msensus	YIKLKR- FVQLKN- GYGKSDAPX-XDXYS 330 GYGKSDQPK-KTGFS GYGFSEASS-KKGFF GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGESSAPPEIEEY GYGESSAPPEIEEY GYGETDAPIHRONY GYNTIDRPSGISDY H- 410 Y-DILGWLIPSTLSS LGQRFGRFLG-LTF H- ES	VRLHYVOT IRMHYVEE 340 SQLACARVER SYNTARIFY SYNTARIFY SYNTARIFY MEVLCKEMV IDLLLVDIK SIDELTGDIR ILTHLVEDIR	GSDDi GPADi GPALI GPA	CPLMLFIHGYP DOVLLMVHGFP - KXXLXGLOWG 360 KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KFYLQGGWG KYFGHWG	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY STATE GILVEYMALE GLIVAYMALE ALLAWHFSIY GMILWHALIAII GLVAWQFAEG AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFF LETONVPLEGGIV. ADGALD LEAA'	ALIDMR A
sh2 sh3 sh3 sh4 sh1 sh2 sh3 sh4 sh1 sh2 sh3 sh3 sh4 sh1 sh2	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGESSAPP	-VRLHYVOTIRMHYVEE -IRMHYVEE -IRMH	GSDDi GPADi GPALI GPA	CPLMLFIHGYP DOVLLMVHGFP - RXXLXGLOWS 360 RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RINLLSHDYG QAVEIGHDWG KCYLIGBDWG KCYLIGBDWG KCYLIGBDWG KCYLIGBDWG KCYLIGBDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHDWG RAIVVAHD RAIVVAHDWG RAIVVAHD RAIVVAHDWG	EFWYSWRFQI EFWYSWRFQI 370 AIITSLITKY SALITY	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFF LETONVPLEGGIV. ADGALD LEAA'	ALIDMR A
sh3 H g1/MEST H 3 4 sh1 sh2 msensus sh3 H g1/MEST H 4 sh1 sh2 msensus sh3 H g1/MEST H 3 4 sh1 sh2 msensus	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEY GYGESSAPPEIEEY GYGESSAPPEIEEY GYGETDAPIHRONY GYNTIDRPSGISDY H		GSDDi GPADi GPALIX-I GPA	CPLMLFIHGYP DOVLLMVHGFP - RXXLXGLOWS 360 RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RYYLQGGWG RYYLGGWG R	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY STATE GILVEYMALE GLIVAYMALE ALLAWHFSIY GMILWHALIAII GLVAWQFAEG AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFF LETONVPLEGGIV. ADGALD LEAA'	ALIDMR 1 XEXXLX 400 NALGTF 2 STLTLL 2 GIFPET 1 FPMMSPL 3 QOYSLB 2 FRIENT 2 RRRIYT 2 FRIENT 480 ENRAL 3 EFFLYNM 1 GIRXX- 480 ENRAL 3 GIRXX- 480 ENRAL 3 GIRXX- 480 ENRAL 3 EFFLYNM 1 GIRXX- 480 ENRAL 3 EFFLYNM 2 GIRXX- 480 ENRAL 3 EFFLYNM 4 GIRXX- 480 ENRAL 3 EFFLYNM 5 GIRXX- 480 ENRAL 3 EFFLYNM 4 ELLYRK 2 MEQHME 5 EXTRE 3 EVT-KI 3 VDS-L 3 ELS-V 3
sh2 sh3 sh3 sh4 sh1 sh2 shsensus sh3 sh4 sh1 sh2 shsensus sh3 sh4 sh1 sh2 sh3 sh4 sh4	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEYC GYGFSDAPRDVDCYT GYMLSDKPKHVDNYS GYNTIDRPSGISDYN H		GSDDi GPADi GPA	CPLMLFIHGYP DOVLLMVHGFP	EFWYSWRFQI EFWYSWRFQI 370 AIITSLITKY SALITY	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFE LAMFRYPYPEFFFE LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFF LAMFRYPYPF	ALIDMR A
eh2 onsensus eh3 eH g1/MEST H i3 i4 eh1 eh2 onsensus eh3 eH i3 i4 eh1 eh2 onsensus eh3 eH i3 i4 eh1 eh2 onsensus	YIKLKR- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGESSAPPELEEY GYGESDAPRDVDCYT GYMLSDKPKHVDNY GYNTDRPSGISDYN H	VRLHYVOT VRHYVEE 340 340 SQLACARVER SIDXLARDIRE SIDXLARDIRE SIDXLARDIRE SIDXLARDIRE SIDXLARDIRE SIDXLARDIRE MEVICKEMV TIDLLLVDIK KIDCLITDIK SIDELITGDIRE NITHLVEDIRE 420 420 SKEIQKTHNP ERDVELLYPV	GSDDi GPRDi GPR	CPLMLFIHGYP DVLLMVHGFP RXXLXGLOWG 350 -KFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG RRINLISHDYG -QAVFIGHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KCILVAHDWG -KVLIGHDWG -KCILVAHDWG -KVLIGHDWG -KCILVAHDWG -KVLIGHDWG -KCILVAHDWG -KVILGHWG -KVILGHW	EFWYSWRFQI EFWYSWRFQI 370 AIITSLITKY SALITY	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFE LAMFRYPYPEFFFE LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFF LAMFRYPYPF	ALIDMR 1
eh2 onsensus eh3 eH g1/MEST H i3 4 4 oh1 oh2 onsensus eh3 eH g1/MEST H i3 4 eh1 eh2 onsensus eh3 eH g1/MEST H i3 i4 eh1 eh2 onsensus eh3 eH g1/MEST H i3 i4 eh1 eh2 onsensus	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFP GFGFSDKPR-PHHYS GYGESSAPPEIEEYC GYGFSDAPRDVDCYT GYMLSDKPKHVDNYS GYNTIDRPSGISDYN H		GSDDI GPADI GPA	CPLMLFIHGYP DOVLLMVHGFP - RXXLXGLOWS 360 RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RFYLQGGWG RAIVAHDWG RAIVAHD RAIVAHDWG	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY SLICTNMAQI GIVAVALIA GIVAWALIA GLVAWALIA GLVAWGFAE AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFE LAMFRYPYPEFFFE LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFF LAMFRYPYPF	ALIDMR A
sh2 sh3 sh3 sh3 sh4 sh1 sh2 sh2 sh5 sh3 sh4 sh1 sh2 sh3 sh3 sh3 sh3 sh3 sh4 sh2 sh3 sh3 sh4 sh2 sh3 sh3 sh4 sh2 sh3 sh3 sh4 sh1 sh2 sh3 sh4 sh2 sh3 sh4 sh2 sh3 sh4 sh2 sh3 sh3 sh4 sh2 sh3 sh4 sh3	YIKLKK- FVQLKN- 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFN GFGFSDKPR-PHHYS GYGESSAPPELEEY GYGPSDAPRDVDCYT GYMLSDKPKHUDNY GYNLSDKPKHUDNY GYNTDRPSGISDYN H	-VRLHYVOTIRMIYVEE -IRMIYVEE -IRMIYVEE -IRMIYVEE -IRMIYVEE	GSDDi GPRDi GPR	CPLMLFIHGYP DOVLLMVHGFP RXXLXGLOWG 350 -KFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLQGDWG -EFYLGGDWG -KAILVAHDWG -CAVPIGHDWG -KCILVAHDWG	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY SLICTNMAQI GIVAVALIA GIVAWALIA GLVAWALIA GLVAWGFAE AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFE LAMFRYPYPEFFFE LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFF LAMFRYPYPF	ALIDMR 1
sh2 msensus sh3 H g1/MEST H 3 4 bh1 sh2 msensus sh3 H g1/MEST H 3 4 4 sh1 sh2 msensus sh3 H g1/MEST H 3 4 4 sh1 sh2 msensus	YIKLKK- FVQLKN- GYGKSDAPX-XDXYS 330 GYGWSDQPK-KTGFS GYGFSEASS-KKGFF GFGFSDKPR-PHHYS GYGESSAPPEIEEYC GYGFSDAPRDVDCYT GYMLSDKPKHVDNYS GYNTTDRPSGISDYN H		GSDDI GPADI GPA	CPLMLFIHGYP DOVLLMVHGFP - KXXLXGLOWS 360	EFWYSWRFQI EFWYSWRFQI EFWYSWRFQI 370 AIITSLLTKY SLICTNMAQI GIVAVALIA GIVAWALIA GLVAWALIA GLVAWGFAE AIVCWRVAMI PX	MEFAD- LEHFKH	XXXXPHPSXF: 390 LIMMALVISNF: ITRSLCLSNG SLNTPFIPAN: VVSGAPMSVY: LEKFFTWTXT: 470 LEKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTNT: EKFSTWTTT EKHLFTSHSTE LEAMFRGSKA DAXLXLETA 550 SLNELYDRTP. FPFELLH-TPF: LEMFNYPEFF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LEAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFFELLH-TPF LAMFRYPYPEFFFELLH-TPF LAMFRYPYPEFFFE LAMFRYPYPEFFFE LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFFF LAMFRYPYPEFF LAMFRYPYPF	ALIDMR 1

Rational of the nomenclature for the new epoxide hydrolases

The family of α/β hydrolase fold epoxide hydrolases in mammals appears to be small, according to our present analysis. Two hitherto known members, mEH and sEH, and two to three new members (EH3 and EH4 and, possibly peg1/MEST, then to be renamed to EH5) result in a low complexity. This made as deviate from two possible nomenclature alternatives and keep the designation most simple by using consecutive numbers.

One of the obvious alternatives would have been to follow the more complex nomenclature system of, e.g. the CYP enzyme superfamily (Nelson (2006) Cytochrome P450 nomenclature, 2004. *Methods Mol Biol* 320: 1-10) that indicates the degree of similarity by sorting the members into subfamilies within families. However, with presently not more than 5 enzymes in total and only one family having more than 1 member, this seemed unnecessary.

We also did not adapt the accurate, complex scheme recently established for epoxide hydrolases from all phylae (van Loo et al. (2006) Diversity and biocatalytic potential of epoxide hydrolases identified by genome analysis. *Applied and Environmental Microbiology* 72: 2905-2917.), according to which EH3 and EH4 belong to group 2, mEH belongs to group 5 and sEH belongs to group 8 of the α/β hydrolase fold epoxide hydrolases, because this would lead to significant confusion within the large community of researchers with an exclusive interest in human/mammalian EHs.

Primer and PCR conditions for the generation of EH3 mutants

pFastBac EH3 served as the template for the mutagenesis PCR. Reaction set up was composed as follows: 10-30 ng plasmid, forward and reverse primer (see Tab. SII), 200 ng each, dNTPs, 250 µM each, and 1 µI Pfu Ultra™ HF ÖÞŒÚ[|ˆ{ ^¦æ^Á in a final volume of 50 µI Pfu Ultra HF reaction buffer (Stratagene, ŠǽR ||ǽŽÔŒÈ

The PCR reaction was started with the initial denaturation at 95 °C for 3 min, followed by 30 cycles of: (1) denaturation for 40 s at 95 °C, (2) annealing for 30 seconds (for individual temperatures used for each primer pair see Tab. SII), and (3) extension for 6 min at 72 °C. The reaction was completed with a final extension at 72 °C for 10 min. Subsequently, the PCR products were treated with Dpn I (one hour at 37 °C) to digest the methylated wild-type parental plasmid. Finally, E.coli DH10B were transformed with 1 μ I of each reaction and plasmids isolated from individual colonies obtained after plating were analyzed for the presence of the desired mutations and the absence of any other sequence alterations. The final constructs obtained this way were further processed to obtain recombinant baculoviruses as described in Materials and Methods.

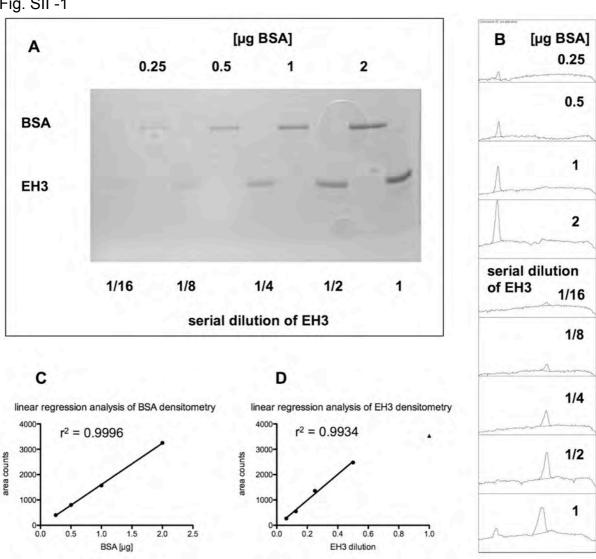
Tab. SII Primer sequences and annealing temperatures

Amino acid substitution	PCR-primers	Annealing temperature [°C]
D173A	forward: GTGGCCCATGCCTGGGGTGCCCT reverse: AGGAGGGCACCCCAGGCATGGG	66
D173N	forward: GTGGCCCATAACTGGGGTGCCCT reverse: AGGAGGGCACCCCAGTTATGGG	62
Y220F	forward: CCGTTCCCACTTCATGTTCCTG reverse: GCTGGAACAGGAACATGAAGTGGG	58
Y280F	forward: CCTCAACTTCTACCGAAACC reverse: CTGAAGAGGTTTCGGTAGAAGTTG	54
Y281F	forward: CCTCAACTACTTCCGAAACC reverse: CTGAAGAGGTTTCGGAAGTAGTTG	54
D307A	forward: GGAGAAGGCCACTTACTTGGAGC reverse: GCTCCAAGTAAGTGGCCTTCTCC	60
D307N	forward: GCTGTGGGGGGAGAAGAACACT reverse: GCTCCAAGTAAGTGTTCTTCTCC	56
H337A	forward: AGGCATAGGGGCTTGGATCCCAC reverse: TGCTCTGTGGGATCCAAGCCCCT	62
H337Q	forward: GCATAGGGCAATGGATCCCACAG reverse: GCTCTGTGGGATCCATTGCCC	60

Immunoquantification of EH3 in insect cell lysates

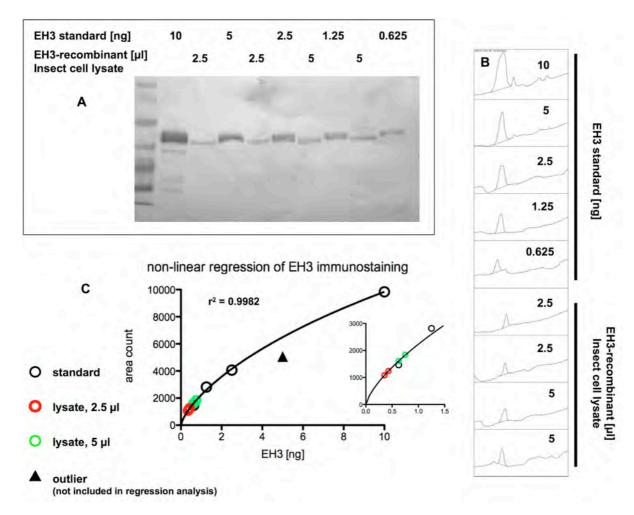
For the calculation of the kinetic properties of EH3, the enzyme was immunoquantified in insect cell lysates. In a first step, recombinant EH3 obtained as inclusion bodies via overexpression in E.coli was washed in Tris-buffered saline (Tris-HCl, 20 mM, NaCl, 100 mM, pH 8.0), inclusion bodies were recovered by brief centrifugation in a benchtop centrifuge and dissolved in high SDS sample buffer (sodium dodecyl sulfate, 5 %, glycerol, 8 %, 2-mercaptoethanol, 3.3 %) to yield a final concentration in the range of $0 - 0.3 \mu g/\mu l$. Ten microliters of 1:1 serial dilutions of the dissolved EH3 protein were analyzed via SDS-PAGE and subsequent Coomassie blue staining, and compared to samples with known, increasing concentrations (0.25 - 2 µg per lane) of a bovine serum albumine that were run along. Gels were photographed with the digital camera of a gel documentation system (GeneFlash®, Syngene, Cambridge, UK) and the band intensities were quantified with the public domain software package ImageJ (NIH, USA), using the routines of the *Analyze/Gels* sub menue (Fig. SlÍ -1). This way, an immunological standard of known EH3 concentration was established.

Fig. SIÍ -1



Using this standard, the recombinant EH3 present in insect cell lysates was quantified by comparative immunoblotting, again using the densitometric features of ImageJ (Fig. SlÍ -2).

Fig. SlÍ -2



The Coomassie-stained polyacrylamid gel (SlÍ -1) and the immunoblot (SlÍ -2) were obtained as digital image (A) and densitometry was performed simulatenously on all lanes of the same image (B). Standard curves were obtained by linear (SlÍ -1) or non-linear (SlÍ -2) regression (C) using *Prism 5* and compared to either dilution series also analyzed by linear regression in the case of Coomassie-stained standard(SlÍ -1 D) or by direct reading from the standard curve in the case of immunoblot analysis (SlÍ -2 C). In the present example, the resulting concentration obtained for EH3 in the analyzed insect cell lysate was 0.15 \pm 0.02 $\mu g/ml$ (mean $\pm SD$).

Structure and inhibitory potency of EH inhibitors employed in the study

IUPAC name (common acronym)	structure	Inhibitory potency
1-(1-acetylpiperidin-4-yl)-3- (4-(trifluoromethoxy)phenyl) urea (TPAU)	F F N N N	+++
12-(3-((3S,5S,7S)- adamantan-1-yl)ureido) dodecanoic acid (AUDA)	Physical Phy	+++
1-cyclohexyl-3-dodecylurea (CDU)	Q H H M M M M M M M M M M M M M M M M M	+++
1-((3 <i>S</i> ,5 <i>S</i> ,7 <i>S</i>)-adamantan-1-yl)-3-(5-(2-(2-ethoxyethoxy) ethoxy)pentyl)urea (AEPU)		++
8-(3-((3 <i>S</i> ,5 <i>S</i> ,7 <i>S</i>)-adamantan- 1-yl)ureido)octanoic acid (AUOA)	ДД Н Н О О О О О О О О О О О О О О О О О	+
4-(((1 <i>R</i> ,4 <i>R</i>)-4-(3-(4- (trifluoromethoxy)phenyl)ur eido)cyclohexyl)oxy)benzoic acid (t-TAUCB)	F F O O O O O O O O O O O O O O O O O O	-
4-(((1S,4S)-4-(3-((3S,5S,7S)-adamantan-1-yl)ureido) cyclohexyl)oxy)benzoic acid (c-AUCB)	Он	-
4-(((1 <i>R</i> ,4 <i>R</i>)-4-(3- ((3 <i>S</i> ,5 <i>S</i> ,7 <i>S</i>)-adamantan-1- yl)ureido)cyclohexyl)oxy)be nzoic acid (t-AUCB)	H H OOH	-
1-((3S,5S,7S)-adamantan-1-yl)-3-cyclohexylurea (ACU)	OH OH	-
(E)-octadec-9-enamide (Elaidamide)	H ₂ N —	-